EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

An advanced Bayesian model for the visual tracking of multiple interacting objects

Carlos R del Blanco*, Fernando Jaureguizar and Narciso García

Abstract

Visual tracking of multiple objects is a key component of many visual-based systems. While there are reliable algorithms for tracking a single object in constrained scenarios, the object tracking is still a challenge in uncontrolled situations involving multiple interacting objects that have a complex dynamics. In this article, a novel Bayesian model for tracking multiple interacting objects in unrestricted situations is proposed. This is accomplished by means of an advanced object dynamic model that predicts possible interactive behaviors, which in turn depend on the inference of potential events of object occlusion. The proposed tracking model can also handle false and missing detections that are typical from visual object detectors operating in uncontrolled scenarios. On the other hand, a Rao-Blackwellization technique has been used to improve the accuracy of the estimated object trajectories, which is a fundamental aspect in the tracking of multiple objects due to its high dimensionality. Excellent results have been obtained using a publicly available database, proving the efficiency of the proposed approach.

Keywords: visual tracking, multiple objects, interacting model, particle filter, Rao-Blackwellization, data association

1 Introduction

Visual object tracking is a fundamental part in many video-based systems such as vehicle navigation, traffic monitoring, human-computer interaction, motion-based recognition, security and surveillance, etc. While there exist reliable algorithms for the tracking of a single object in constrained scenarios, the object tracking is still a challenge in uncontrolled situations involving multiple objects with complex dynamics. The main problem is that object detectors produce a set of unlabeled and unordered detections, whose correspondence with the tracked objects is unknown. The estimation of this correspondence, called the data association problem, is of paramount importance for the proper estimation of the object trajectories. In addition, visual object detectors can produce false and missing detections as consequence of object appearance changes, illumination variations, occlusions, and scene structures similar to the objects of interest (also called clutter). This fact makes more complex the estimation of the true correspondence between detections and objects. Another important issue related to the data association is the computational cost, since it grows exponentially with the number of objects.

To alleviate the data association problem, the tracking also relies on the prior knowledge about the object dynamics, which constrains the feasible associations between detections and objects. Nonetheless, the modeling of the object dynamics can be a very difficult task, especially in situations in which the objects undergo complex interactions.

Besides, the estimation of the object trajectories can be quite inaccurate in situations involving many objects due to the high dimensionality of the resulting tracking problem, which is called the curse of dimensionality [1].

In this article, an efficient Bayesian tracking framework for multiple interacting objects in complex situations is proposed. Complex object interactions are simulated by means of a novel dynamic model that uses potential events of object occlusions to predict different object behaviors. This interacting dynamic model allows to appropriately estimate a set of data association hypotheses that are used for the estimation of the object trajectories. On the other hand, a Rao-Blackwellization strategy [2] has been used to derive an approximation of the posterior distribution over the object trajectories,

* Correspondence: cda@gti.ssr.upm.es
Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, Madrid, 28040, Spain

which allows to achieve accurate estimates in spite of the high dimensionality.

The organization of the article is as follows. The state of the art is presented in Section 2. The description of the tracking model for interacting objects is described in Section 3. The inference method used to estimate the object trajectories from the given tracking model is presented in Sections 4, 5, and 6. Results are shown in Section 7, and lastly, conclusions are drawn in Section 8.

2 State of the art

Many strategies have been proposed in the scientific literature to solve the data association problem. The simplest one is the global nearest neighbor algorithm [3], also known as the 2D assignment algorithm, which computes a single association between detections and objects. However, this approach discards many feasible associations. On the other hand, the multiple hypotheses tracker (MHT) [4,5] attempts to compute all the possible associations along the time. However, the number of associations grows exponentially over time, and consequently the computational cost becomes prohibitive. Therefore, a trade-off between computational efficiency and handling of multiple association hypotheses is needed. In this respect, one of the most popular methods is the joint probabilistic data association filter (JPDAF) [6,7], which performs a soft association between detections and objects. This consists in combining all the detections with all the objects, which prunes away many unfeasible hypotheses, but also restricts the data association distribution to be Gaussian. Subsequent works [8,9] have tried to overcome this limitation using a mixture of Gaussians to model the data association distribution. However, heuristic techniques are necessary to prune the number of components and make the algorithm computationally manageable. The probabilistic multiple hypotheses tracker (PMHT) [10,11] assumes that the data association is an independent process to overcome the problems with the pruning. Nevertheless, the performance is similar to that of the JPDAF, although the computational cost is higher.

The data association problem has been also addressed with particle filtering techniques. These allow to deal with arbitrary data association distributions in a natural way, establishing a compromise between the computational cost and the accuracy in the estimation. In practice, the performance of the particle filtering techniques depends on the ability to correctly sample association hypotheses from a proposal distribution. In [12], a Gibbs sampler is used to sample the data association hypotheses, while in [13,14] a strategy based on a Markov Chain Monte Carlo (MCMC) is followed. The main problem with these samplers is that they are iterative methods that need an unknown number of iterations to converge. This fact can make them inappropriate for online applications. Some works [15-17] overcome this limitation by designing an efficient and non-iterative proposal distribution that depends on the specific characteristics of the tracking system. An additional problem is that the accuracy of the estimated object trajectories can be very poor due to the high dimensionality of the tracking problem. In [18], a variance reduction technique called Rao-Blackwellization has been used to improve the accuracy.

A random finite set (RFS) approach can be used as an alternative to data association methods, which treats the collection of objects and detections as finite sets. However, the computation of the posterior of a RFS is intractable in general, and therefore the use of approximations is required. In [19], a probability hypothesis density (PHD) filter is used in the context of visual tracking, which approximates the full posterior distribution by its first-order moment. The cardinalized PHD (CPHD) filter [20] is a variation of the PHD that is able to propagate the entire probability distribution on the number of objects. In [21], a closed form for the posterior distribution is derived assuming that the image regions that are influenced by individual states do not overlap.

One common limitation of the previous works is their limitation to track interacting objects. They cannot manage complex interactions involving trajectory changes and occlusions, since the assumption that the objects move independently does not hold. Part of the problem comes from the fact that these techniques were developed for radar and sonar applications, in which the dynamics of the target objects have certain physical restrictions that prevent the existence of the complex interactions that can occur in visual tracking. On the other hand, tracked objects are usually considered as point targets [22]. Therefore, occlusion events between tracked objects are not as problematic as in the field of visual tracking, wherein they are one of the main sources of tracking errors. Some works have proposed specific strategies to deal with the problems that arise in visual tracking. In [23,24] data association hypotheses are computed using a sampling technique that is able to handle split and merged detections. These type of detections are typical from background subtraction techniques [25], which are used to detect moving objects in video sequences. In [26], an approach for handling object interactions involving occlusions and changes in trajectories is proposed. It creates virtual detections of possible occluded objects to cope with the changes in trajectories during the occlusions. However, tracking errors can appear when a virtual detection is associated to an object that is actually not occluded. In this article, a novel Bayesian approach that explicitly models the

occlusion phenomenon and the object interactions has been developed, which is able to reliably track complex interacting objects whose trajectories change during occlusions.

3 Bayesian tracking model for multiple interacting objects

The aim is to track several interacting objects from a static camera. From a Bayesian perspective, this is accomplished by estimating the posterior probability density function (pdf) over the object trajectories $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ using a sequence of noisy detections and the prior information about the object dynamics. This probability contains all the required information to compute an optimum estimate of the object trajectories at each time step. The information about the object trajectories at the time step t is represented by the state vector

$$\mathbf{x}_t = \{\mathbf{x}_{t,i}|i = 1, \ldots, N_{\text{obj}}\}, \tag{1}$$

where each component contains the 2D position and velocity of a tracked object. The number of tracked objects N_{obj} is variable, but it is assumed that entrances and exits of objects in the scene are known. This allows to focus on the modeling of object interactions.

The sequence of available detections until the current time step is represented by $\mathbf{z}_{1:t} = \{\mathbf{z}_1, ..., \mathbf{z}_t\}$, where $\mathbf{z}_t = \{\mathbf{z}_{t,j}|j = 1, ..., N_{ms}\}$ contains the set of detections at the current time step t. The number of detections N_{ms} can vary at each time step. Each detection $\mathbf{z}_{t,j}$ contains the position of a potential object, and a confidence value related to the quality of the detection. Detections are obtained from each frame by means of a set of object detectors, where each detector is specialized in one specific type or category of object. Detections have associated an object category identifier according to the object detector that created them. In addition, some of the computed detections can be false alarms due to the clutter, and also there can be objects without any detection, called missing detections, as consequence of occlusions and changes in the object appearance and illumination.

The detections at each time step are unordered and partially unlabeled. The object category of a detection is known, but its correspondence with a specific object inside a category is unknown. Consequently, the data association between detections and objects has to be estimated. The data association is modeled by the random variable

$$\mathbf{a}_t = \{a_{t,j}|j = 1, \ldots, N_{ms}\}, \tag{2}$$

where the component $a_{t,j}$ specifies the association of the jth detection $\mathbf{z}_{t,j}$. A detection can be associated to one object or to the clutter, indicating in this last case that it is a false alarm. The association of the jth detection with the ith object is expressed as $\mathbf{a}_{t,j} = i$, while the association with the clutter is expressed as $\mathbf{a}_{t,j} = 0$. Figure 1 illustrates the data association process between detections and objects.

The prior knowledge about the object dynamics is used to improve the estimation of the object state as well as to reduce the ambiguity in the data association estimation. The proposed interacting dynamic model predicts different object behaviors depending on the events of occlusions. This fact implies that the object occlusions must be estimated. The object occlusions are modeled by the random variable

$$\mathbf{o}_t = \{\mathbf{o}_{t,i}|i = 1, \ldots, N_{\text{obj}}\}, \tag{3}$$

where each component stores the occlusion information of one object. To express that the ith object is occluded by the lth object, $\mathbf{o}_{t,i} = l$ is written. And, if the object is not occluded, it is expressed as $\mathbf{o}_{t,i} = 0$.

The variables $\mathbf{a}_t$ and $\mathbf{o}_t$ are necessary to estimate the posterior pdf over the object trajectories. This fact can be observed in the graphical model of Figure 2, which shows the probabilistic dependencies among the different random variables involved in the tracking task. According to this, the posterior pdf is expressed as

$$p(\mathbf{x}_t|\mathbf{z}_{1:t}) = \sum_{\mathbf{a}_t}\sum_{\mathbf{o}_t} p(\mathbf{x}_t, \mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t}), \tag{4}$$

where the joint posterior pdf can be recursively expressed using the Bayes' theorem as

$$p(\mathbf{x}_t, \mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t}) = \frac{p(\mathbf{z}_t|\mathbf{z}_{1:t-1}, \mathbf{x}_t, \mathbf{a}_t, \mathbf{o}_t)p(\mathbf{x}_t, \mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t-1})}{p(\mathbf{z}_t|\mathbf{z}_{1:t-1})}, \tag{5}$$

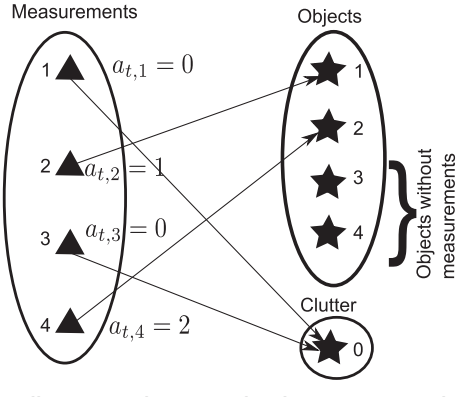

Figure 1 Illustration depicting the data association between detections and objects.

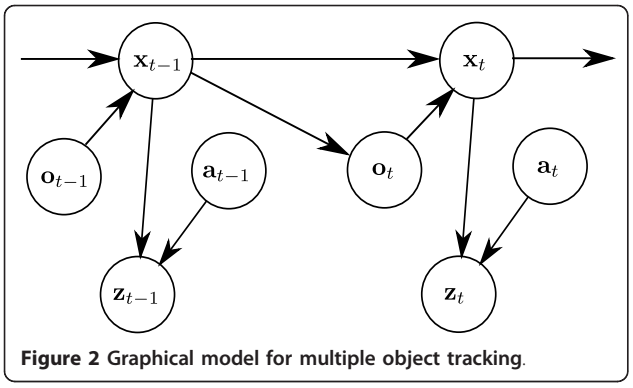

Figure 2 Graphical model for multiple object tracking.

where the probability term in the denominator is just a normalization constant, and the other terms as explained as follows.

The term $p(\mathbf{x}_t, \mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t-1})$ is the prior pdf that predicts the evolution of $\{\mathbf{x}_t, \mathbf{a}_t, \mathbf{o}_t\}$ between consecutive time steps using the joint posterior pdf at the previous time step $p(\mathbf{x}_{t-1}, \mathbf{a}_{t-1}, \mathbf{o}_{t-1}|\mathbf{z}_{1:t-1})$

$$\begin{aligned} &p(\mathbf{x}_t, \mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t-1}) \\ &= \int \sum_{\mathbf{a}_{t-1}} \sum_{\mathbf{o}_{t-1}} p(\mathbf{x}_t, \mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t-1}, \mathbf{x}_{t-1}, \mathbf{a}_{t-1}, \mathbf{o}_{t-1}) \\ &.p(\mathbf{x}_{t-1}, \mathbf{a}_{t-1}, \mathbf{o}_{t-1}|\mathbf{z}_{1:t-1}) d\mathbf{x}_{t-1}. \end{aligned} \quad (6)$$

The transition term $p(\mathbf{x}_t, \mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t-1}, \mathbf{x}_{t-1}, \mathbf{a}_{t-1}, \mathbf{o}_{t-1})$ can be factorized as

$$\begin{aligned} &p(\mathbf{x}_t, \mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t-1}, \mathbf{x}_{t-1}, \mathbf{a}_{t-1}, \mathbf{o}_{t-1}) \\ &= p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{o}_t) p(\mathbf{a}_t) p(\mathbf{o}_t|\mathbf{x}_{t-1}), \end{aligned} \quad (7)$$

taking into account the conditional independence properties of the involved variables (see [27,28] for an explanation of how to derive and apply the conditional independence properties given a graphical model). From now on, the conditional independence properties will be applied whenever possible to simplify probabilities expressions. These properties expresses three different characteristics of the tracking problem: first, $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{o}_t)$, that models the dynamics of interacting objects, depends only on the previous object positions and possible occlusions; second, since the detections are unordered, previous data associations and object positions are useless for the prediction of the current data association $p(\mathbf{a}_t)$; and last, $p(\mathbf{o}_t|\mathbf{x}_{t-1})$, that models the object occlusions, depends only on the previous object positions.

Using the new set of available detections at the current time, the prediction on $\{\mathbf{x}_t, \mathbf{a}_t, \mathbf{o}_t\}$ is rectified by the likelihood term of Equation 5, which can be simplified as

$$p(\mathbf{z}_t|\mathbf{z}_{1:t-1}, \mathbf{x}_t, \mathbf{a}_t, \mathbf{o}_t) = p(\mathbf{z}_t|\mathbf{x}_t, \mathbf{a}_t). \quad (8)$$

This expression reflects the fact that the data association between detections and objects is necessary for estimating the object trajectories.

Lastly, the object trajectories at the current time step are obtained by computing the maximum a posteriori (MAP) estimation of $p(\mathbf{x}_t|\mathbf{z}_{1:t})$.

However, $p(\mathbf{x}_t, \mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t})$ cannot be analytically solved, and therefore neither can $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ be. This problem arises from the fact that some of the stochastic processes involved in the multiple object tracking model are nonlinear or/and non-Gaussian [29]. To overcome this problem, an approximate inference technique is introduced in the next section that allows to obtain an accurate suboptimal solution.

4 Approximate inference based on a Rao-Blackwellized particle filtering

The variance reduction technique Rao-Blackwellization has been used to accurately approximate $p(\mathbf{x}_t, \mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t})$ This technique assumes that the random variables have a special structure that allows to analytically marginalize out some of the variables conditioned to the rest ones, improving the estimation in high dimensional problems.

In the proposed Bayesian tracking model, the object state $\mathbf{x}_t$ can be marginalized out conditioned to $\{\mathbf{a}_t, \mathbf{o}_t\}$. Thus, the Rao-Blackwellization technique can be applied to express the joint posterior pdf as

$$\begin{aligned} &p(\mathbf{x}_t, \mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t}) \\ &= p(\mathbf{x}_t|\mathbf{z}_{1:t}, \mathbf{a}_t, \mathbf{o}_t) p(\mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t}), \end{aligned} \quad (9)$$

where $p(\mathbf{x}_t|\mathbf{z}_{1:t}, \mathbf{a}_t, \mathbf{o}_t)$ is assumed to be conditionally linear Gaussian, and therefore with an analytical expression known as the Kalman filter. This assumption arises from the fact that the object dynamics can be acceptably simulated by a constant velocity model with Gaussian perturbations if the object occlusions and the data association are known. That is, if the main sources of nonlinearity and multimodality in the tracking problem are known. Section 5 derives the expression of $p(\mathbf{x}_t|\mathbf{z}_{1:t}, \mathbf{a}_t, \mathbf{o}_t)$ using a dynamic model for interacting objects.

The other probability term in Equation 9 can be expressed using the Bayes' theorem as

$$p(\mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t}) = \frac{p(\mathbf{z}_t|\mathbf{z}_{1:t-1}, \mathbf{a}_t, \mathbf{o}_t) p(\mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t-1})}{p(\mathbf{z}_t|\mathbf{z}_{1:t-1})}. \quad (10)$$

The prior term $p(\mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t-1})$ can be recursively expressed as

$$\begin{aligned} p(\mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t-1}) = &\sum_{\mathbf{a}_{t-1}} \sum_{\mathbf{o}_{t-1}} p(\mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t-1}, \mathbf{a}_{t-1}, \mathbf{o}_{t-1}) \\ &.p(\mathbf{a}_{t-1}, \mathbf{o}_{t-1}|\mathbf{z}_{1:t-1}), \end{aligned} \quad (11)$$

where the transition term can be factorized and simplified as

$$p(\mathbf{a}_t, \mathbf{o}_t|\mathbf{z}_{1:t-1}, \mathbf{a}_{t-1}, \mathbf{o}_{t-1}) = p(\mathbf{a}_t)p(\mathbf{o}_t|\mathbf{z}_{1:t-1}, \mathbf{a}_{t-1}, \mathbf{o}_{t-1}). \quad (12)$$

The term $p(\mathbf{a}_t)$ is the prior pdf over the data association and is used to restrict the possible associations between detections and objects. The first restriction establishes that one detection can be only associated with one object or to the clutter, since the region from which was extracted the detection can only belong to one object due to the occlusion phenomenon. The second restriction imposes that one object can be associated at most with one detection, although the clutter can be associated with several detections. This restriction results from the characteristics of the object detector, which does not allow split detections. The last restriction states that, given a group of detections that share common image regions, only one of them can be associated with an object, while the rest are associated to the clutter. This phenomenon happens because an image region could be potentially part of several object instances, and it is not possible to determine the true one. Figure 3a illustrates the first restriction where there are two objects partially occluded and only one detection. This restriction avoids that the detection can be associated to both objects. Figure 3b shows the second restriction where there are only one object and two detections. This restriction ensures that only one detection can be associated with the object, whereas the other is associated with the clutter. Figure 3c illustrates the third restriction where there are two objects partially occluded and three detections. Since one of the objects is too occluded, only one detection should be ideally generated. But, two more are generated from the combination of image regions belonging to both objects.

Mathematically, $p(\mathbf{a}_t)$ is expressed as

$$p(\mathbf{a}_t) = \prod_{j=1}^{N_{ms}} p(\mathbf{a}_{t,j}|\mathbf{a}_{t,1}, \ldots, \mathbf{a}_{t,j-1}), \quad (13)$$

where one association depends on the previous computed associations. If one detection fulfills the second and third restrictions, the object association probability is $p(\mathbf{a}_{t,j} = i|\mathbf{a}_{t,1}, ..., \mathbf{a}_{t,j-1}) = p^{\text{obj}}$ that expresses the prior probability that one detection is associated with one object. In the same conditions, the clutter association probability is $p(\mathbf{a}_{t,j} = 0|\mathbf{a}_{t,1}, ..., \mathbf{a}_{t,j-1}) = p^{\text{clu}}$. If any of the restrictions is not fulfilled, the detection is associated to the clutter.

The other term in Equation 12 can be factorized and simplified as

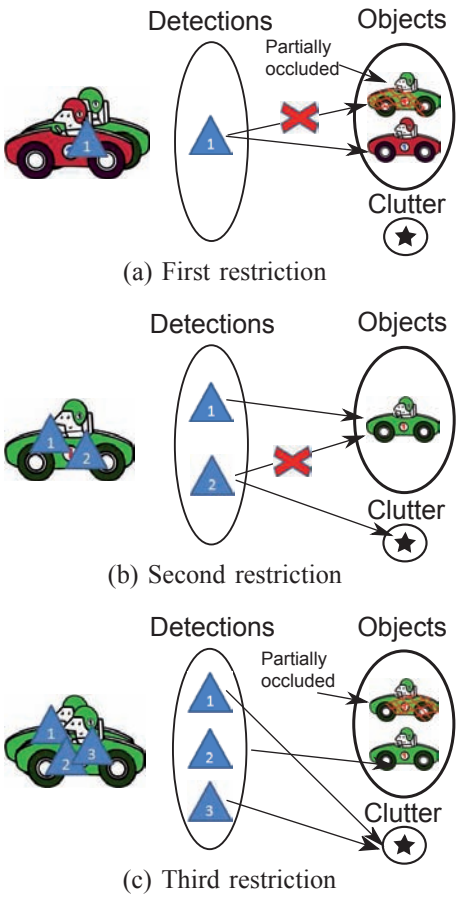

Figure 3 Data association restrictions.

$$p(\mathbf{o}_t|\mathbf{z}_{1:t-1}, \mathbf{a}_{t-1}, \mathbf{o}_{t-1}) = \int p(\mathbf{o}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{a}_{t-1}, \mathbf{o}_{t-1})d\mathbf{x}_{t-1}, \quad (14)$$

where $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}, \mathbf{a}_{t-1}, \mathbf{o}_{t-1})$ is the conditional posterior pdf over the object trajectories in the previous time step, and the term $p(\mathbf{o}_t|\mathbf{x}_{t-1})$ models the occlusion phenomenon among objects. The occlusion model considers that two or more objects are involved in an occlusion if they are enough close each other. Also, some restrictions are imposed. In an occlusion, only one object is considered to be in the foreground, while the rest are occluded behind it. This means that an occluding object cannot be occluded by anyone, and that an occluded object cannot occlude others. Mathematically, this is formulated as

$$p(\mathbf{o}_t|\mathbf{x}_{t-1}) = \prod_{i=1}^{N_{\text{obj}}} p(\mathbf{o}_{t,i}|\mathbf{x}_{t-1}, \mathbf{o}_{t,1}, \ldots, \mathbf{o}_{t,i-1}), \quad (15)$$

where an occlusion event depends on the previous computed occlusions. The probability that one object is

occluded by another, providing that both objects have not been involved in previous occlusion events, is expressed by a Gaussian function that depends on the distance between the two considered objects. And in the same conditions, the probability that it is not occluded is determined by the probability density d^{vis}. In the case that any of the considered objects have been involved in previous occlusion events, the occlusion restrictions are applied to avoid non-realistic situations.

The likelihood term in Equation 10 models the data association process. It can be decomposed and simplified as

$$p(\mathbf{z}_t|\mathbf{z}_{1:t-1},\ \mathbf{a}_t,\ \mathbf{o}_t) = \int p(\mathbf{z}_t|\mathbf{a}_t,\ \mathbf{x}_t)p(\mathbf{x}_t|\mathbf{z}_{1:t-1},\mathbf{o}_t)d\mathbf{x}_t, \tag{16}$$

where $p(\mathbf{x}_t|\mathbf{z}_{1:t\text{-}1},\ \mathbf{o}_t)$ is the prior pdf involved in the conditional Kalman filter used to compute $p(\mathbf{x}_t|\mathbf{z}_{1:t},\ \mathbf{a}_t,\ \mathbf{o}_t)$, and the other term estimates the data association between detections and objects as

$$p(\mathbf{z}_t|\mathbf{x}_t,\ \mathbf{a}_t) = \prod_{j=1}^{N_{ms}} p(\mathbf{z}_{t,j}|\mathbf{x}_t,\ \mathbf{a}_{t,j}). \tag{17}$$

Each factor computes the association likelihood of one detection as

$$p(\mathbf{z}_{t,j}|\mathbf{x}_t,\ \mathbf{a}_{t,j}) = \begin{cases} N(\mathbf{r}^z_{t,j};\mathbf{r}^x_{t,i},\ \mathbf{\Sigma}^{lh}) & \text{if object association,} \\ d^{\text{clu}} & \text{if clutter association,} \end{cases} \tag{18}$$

where $i \in \{1,\ldots, N_{\text{obj}}\}$, $\mathbf{r}^z_{t,j}$ and $\mathbf{r}^x_{t,i}$ are the positional information of the detection and the object, respectively, d^{clu} is the clutter probability density, and $\mathbf{\Sigma}^{lh}$ is the covariance matrix of the Gaussian function. The previous expression is only applicable between detections and objects of the same category, since the object association probability is zero otherwise.

The last probability term $p(\mathbf{z}_t|\mathbf{z}_{1:t\text{-}1})$ in Equation 10 is just a normalization constant.

As occurred with $p(\mathbf{x}_t,\ \mathbf{a}_t,\ \mathbf{o}_t|\mathbf{z}_{1:t})$, the posterior pdf $p(\mathbf{a}_t,\ \mathbf{o}_t|\mathbf{z}_{1:t})$ has not analytical form. To overcome this problem, an approximate inference method based on a particle filtering framework is used to obtain a suboptimal solution, which is described in Section 6.

5 Conditional Kalman filtering of object trajectories

The Kalman filter recursively computes $p(\mathbf{x}_t|\mathbf{z}_{1:t},\ \mathbf{a}_t,\ \mathbf{o}_t)$ in two steps: prediction and update. The prediction step estimates the object trajectories at the current time step according to a dynamic model for interacting objects. This model considers that an interacting behavior mainly occurs when two or more objects are involved in an occlusion event. In case of interaction, one object remains totally or partially occluded behind the occluding object until the interaction ends. This behavior simulates a situation where the occluded object seems to be following the occluding one, changing its trajectory. Another possibility is that the occluded object is not interacting with anyone. In this case, the occluded object keeps its trajectory constant according to a piecewise constant velocity model. Since a priori it is not possible to know if an object is interacting or not in the presence of an occlusion, both hypotheses are propagated along the time. When the occlusion event has ended and there are new detections, these are used to determine which hypothesis was the correct. On the other hand, objects that are not involved in an occlusion move independently according to a piecewise constant velocity model. This approach is very efficient since detections are used to rectify object trajectories, being able to locally approximate nonlinear behaviors. Figure 4 illustrates the previous kinds of situations that the interacting dynamic model can handle.

According to the previous interacting dynamic model, and noting that $\mathbf{x}_t$ is conditionally independent of $\mathbf{a}_t$, the prediction of the object trajectories is expressed by the multivariate Gaussian function

$$p(\mathbf{x}_t|\mathbf{z}_{1:t-1},\mathbf{a}_t,\mathbf{o}_t) = p(\mathbf{x}_t|\mathbf{z}_{1:t-1},\mathbf{o}_t) = N\left(\mathbf{x}_t;\hat{\mu}_t,\hat{\mathbf{\Sigma}}_t\right), \tag{19}$$

where $\hat{\mu}_t$ is the mean, and $\hat{\mathbf{\Sigma}}_t$ is the covariance matrix. If the ith object is not occluded, determined by $\mathbf{o}_{t,\ i} = 0$, its mean is computed by $\hat{\mu}_{t,i} = \mathbf{A}\mu_{t-1,i,}$ where $\mathbf{A}$ is a

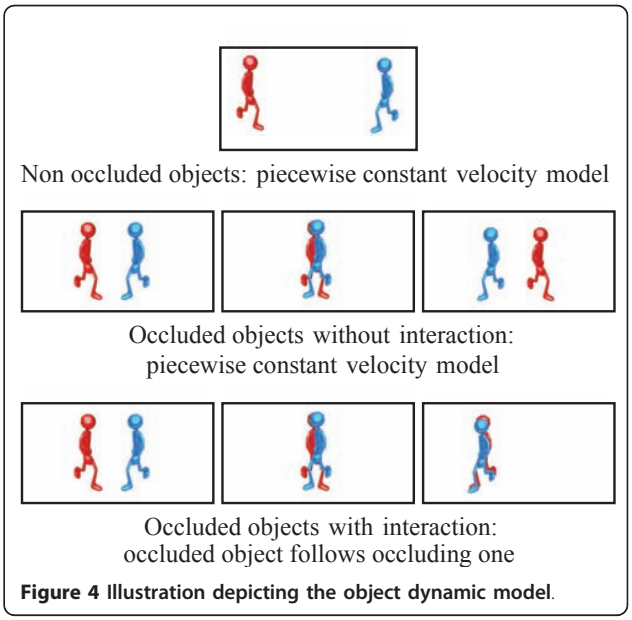

Figure 4 Illustration depicting the object dynamic model.

matrix simulating a constant velocity model. In the case that the object is occluded, determined by $\mathbf{o}_{t,\,i} = l$, there are two different hypotheses

$$\hat{\mu}_{t,i} = \begin{cases} \mathbf{A}\frac{\mu_{t-1,i}+\mu_{t-1,l}}{2} & \text{if interaction,} \\ \mathbf{A}\mu_{t-1,i} & \text{if not interaction,} \end{cases} \tag{20}$$

depending if the object is assumed to undergo an interaction or not. The event of interaction is managed by a Bernoulli distribution, whose parameter can be adjusted according to the expected number of interactions per occlusion.

The covariance matrix $\hat{\mathbf{\Sigma}}_t$ is computed using the standard equations of the Kalman filter, taking into account that the prior covariance for occluded objects should be higher than that for non-occluded ones, since the uncertainty in the trajectory of an occluded object is usually higher.

The second step uses the set of available detections at the current time step to update the previous prediction

$$p(\mathbf{x}_t|\mathbf{z}_{1:t},\ \mathbf{a}_t,\ \mathbf{o}_t) = N(\mathbf{x}_t; \mu_t,\ \mathbf{\Sigma}_t), \tag{21}$$

where the parameters of the Gaussian function are obtained using the standard expressions of the Kalman filter. The update step only is applied to those objects that have associated a detection, determined by $\mathbf{a}_{t,\,j} = i$; $i \in \{1, ..., N_{\text{obj}}\}$.

6 Ancestral particle filtering of data association and object occlusions

The posterior pdf on $\{\mathbf{a}_t, \mathbf{o}_t\}$ is simulated by a set of N_{sam} unweighted samples, also called particles, as

$$p(\mathbf{a}_t,\ \mathbf{o}_t|\mathbf{z}_{1:\,t}) = \sum_{k=1}^{N_{\text{sam}}} \delta\left(\mathbf{a}_t - \mathbf{a}_t^k,\ \mathbf{o}_t - \mathbf{o}_t^k\right), \tag{22}$$

where $\delta(x)$ is a Kronecker delta function, and $\left\{\mathbf{a}_t^k,\ \mathbf{o}_t^k | k = 1, \ldots,\ N_{\text{sam}}\right\}$ are the samples, which are drawn from

$$\begin{aligned} p(\mathbf{a}_t,\ \mathbf{o}_t|\mathbf{z}_{1:t}) &\propto p(\mathbf{z}_t|\mathbf{z}_{1:t-1},\ \mathbf{a}_t,\ \mathbf{o}_t) \\ &\cdot \sum_{k=1}^{N_{\text{sam}}} p(\mathbf{a}_t,\ \mathbf{o}_t|\mathbf{z}_{1:t-1},\ \mathbf{a}_{t-1}^k, \mathbf{o}_{t-1}^k), \end{aligned} \tag{23}$$

where the sampled-based approximation of the posterior pdf in the previous time step has been used. All the probability terms have been already defined in previous sections, therefore substituting their expressions

$$\begin{aligned} p(\mathbf{a}_t,\ \mathbf{o}_t|\mathbf{z}_{1:t}) &\propto p(\mathbf{a}_t) \int p(\mathbf{z}_t|\mathbf{a}_t,\ \mathbf{x}_t) p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{o}_t) d\mathbf{x}_t \\ &\cdot \sum_{k=1}^{N_{sam}} \int p(\mathbf{o}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1},\ \mathbf{a}_{t-1}^k, \mathbf{o}_{t-1}^k) d\mathbf{x}_{t-1}. \end{aligned} \tag{24}$$

The process to draw samples from the previous probability is based on a hierarchical Monte Carlo technique, called ancestral sampling [30]. This technique hierarchically draws samples from the random variables according to their conditional dependencies. Thus, the process to obtain a new sample starts by drawing a sample $\{\mathbf{a}_{t-1}^k, \mathbf{o}_{t-1}^k\}$ from the sample-based approximation of $p(\mathbf{a}_{t\text{-}1},\ \mathbf{o}_{t\text{-}1}|\mathbf{z}_{1:t\text{-}1})$ computed in the previous time step. Conditioned on the previous sample, a sample $\mathbf{o}_t^k$ is drawn from

$$\mathbf{o}_t^k \sim \int p(\mathbf{o}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1},\ \mathbf{a}_{t-1}^k, \mathbf{o}_{t-1}^k) d\mathbf{x}_{t-1}. \tag{25}$$

Since the previous integral has not analytical form, a suboptimal solution is computed. This consists in approximating the Gaussian $p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1},\ \mathbf{a}_{t-1}^k, \mathbf{o}_{t-1}^k)$ by its mean, obtaining

$$\mathbf{o}_t^k \sim p(\mathbf{o}_t|\mu_{t-1}), \tag{26}$$

which is a discrete probability defined in Section 4.

Lastly, a data association sample is drawn from

$$\mathbf{a}_t^k p(\mathbf{a}_t) \int p(\mathbf{z}_t|\mathbf{x}_t,\ \mathbf{a}_t) p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{o}_t^k) d\mathbf{x}_t \tag{27}$$

conditioned to the rest of sampled variables. The computation of the integral is based on the fact that the integral of any function $f(x)$ proportional to a Gaussian is equal to maximum of that function $f(x)^*$ times a proportionality constant [24]. In this case, $p\left(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{o}_t^k\right)$ is Gaussian since it is the prediction step of the Kalman filter, and the expression of $p(\mathbf{z}_t|\mathbf{x}_t, \mathbf{a}_t)$ is proportional to a Gaussian function. And as the product of Gaussian functions is another Gaussian function, the above integral can be computed as

$$f(\mathbf{x}_t; \mathbf{a}_t) = p(\mathbf{z}_t|\mathbf{x}_t,\ \mathbf{a}_t) p(\mathbf{x}_t|\mathbf{z}_{1:t-1}, \mathbf{o}_t^k), \tag{28}$$

$$\int f(\mathbf{x}_t; \mathbf{a}_t) d\mathbf{x}_t = \sqrt{det(2\pi\,\Sigma_f)} f(\mathbf{x}_t; \mathbf{a}_t)^*, \tag{29}$$

where $\mathbf{a}_t$ acts as a parameter of $f(\mathbf{x}_t; \mathbf{a}_t)$, $det()$ is the determinant function, and Σ_f is the covariance matrix of $f(\mathbf{x}_t; \mathbf{a}_t)$.

As a result, data association samples are drawn from

$$\mathbf{a}_t^k \sim p(\mathbf{a}_t) \sqrt{det(2\pi\,\Sigma_f)} f(\mathbf{x}_t; \mathbf{a}_t)^*, \tag{30}$$

where all the involved probability terms are discrete, and whose mathematical expressions are defined in Sections 4 and 5.

7 Results

The proposed Bayesian tracking model for interacting objects has been evaluated using the public database 'VS-PETS 2003' [31], which contains sequences of a football match. Given the great number and variety of player interactions, this dataset is very suitable for testing purposes.

Two different object detectors [26] are used to detect the players of each team, which characterize each object category by means of its color distribution. Although these detectors are not very complex, they are suitable for the detection of players in the considered dataset. Nonetheless, whatever visual object detector can be used with the presented tracking algorithm provided that at least positional information is given. In this sense, the use of more complex detectors would increase the tracking performance. Figures 5 and 6 show the output of every detector for an image of the dataset. Notice that there are missing and false detections due to object occlusions and clutter.

Figure 7 shows a simple cross between two rival players, who keep their trajectories along the occlusion event. The first row shows the original frames with a blue square that encloses the players involved in the simple cross. The second row shows the image regions inside the previous blue squares and the object detections marked with crosses. In the last row, the computed tracked objects have been enclosed in rectangles and labeled with identifiers. Since the objects belong to different categories, the data association is simpler because the detections can be only associated to objects of the same category. A consequence is that the marginal posterior pdfs of the trajectories of the involved objects are unimodal rather than multimodal. This fact can be observed in Figure 8, where the samples represent the means of a mixture of Gaussians that approximate every marginal posterior pdf.

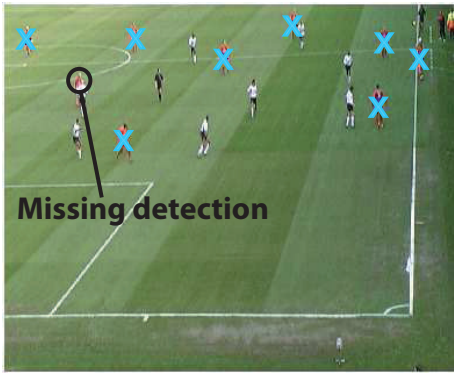

Figure 5 Detected players of the read team.

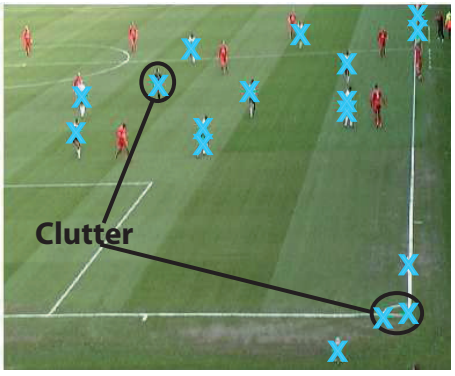

Figure 6 Detected players of the black and white team.

In Figure 9, a complex cross involving three players, two of them from the same team, is shown. In this case, the object trajectories change their direction during the occlusion event. This situation is more complex than a simplex cross since there are several feasible hypotheses for the object dynamics and for the data association. The presented tracking model achieves to successfully track the objects because it is able to compute and manage several hypotheses of object behaviors and data association. In this case, the marginal posterior pdfs of the involved object trajectories are multimodal, as it can be observed in Figure 10.

Figure 11 shows an overtaking action involving three players, two of them belonging to the same team. In this situation, the object trajectories keep their direction during the occlusion like in a simple cross. But, the duration of the occlusion is usually much longer than that for a simple cross. This fact implies more missing detections and a higher uncertainty in the object behavior, and consequently a greater complexity. This leads to multimodal marginal posterior pdfs, as shown in Figure 12.

The proposed tracking algorithm has been compared with the Rao-Blackwellized Monte Carlo data association (RBMCDA) method [18], a state-of-the-art tracking algorithm for multiple objects. Its main characteristics are the ability to handle false and missing detections, and the use of the Rao-Blackwellization technique to achieve accurate estimation in high dimensional state space. The main difference with the algorithm proposed in this article is the lack of an interacting model, which limits its ability to handle object interactions.

Table 1 shows the tracking results for both algorithms, the RBMCDA method and the one presented in this article, which will be called by analogy interacting Rao-Blackwellized Monte Carlo data association (IRBMCDA) method. The results show the number of tracking errors

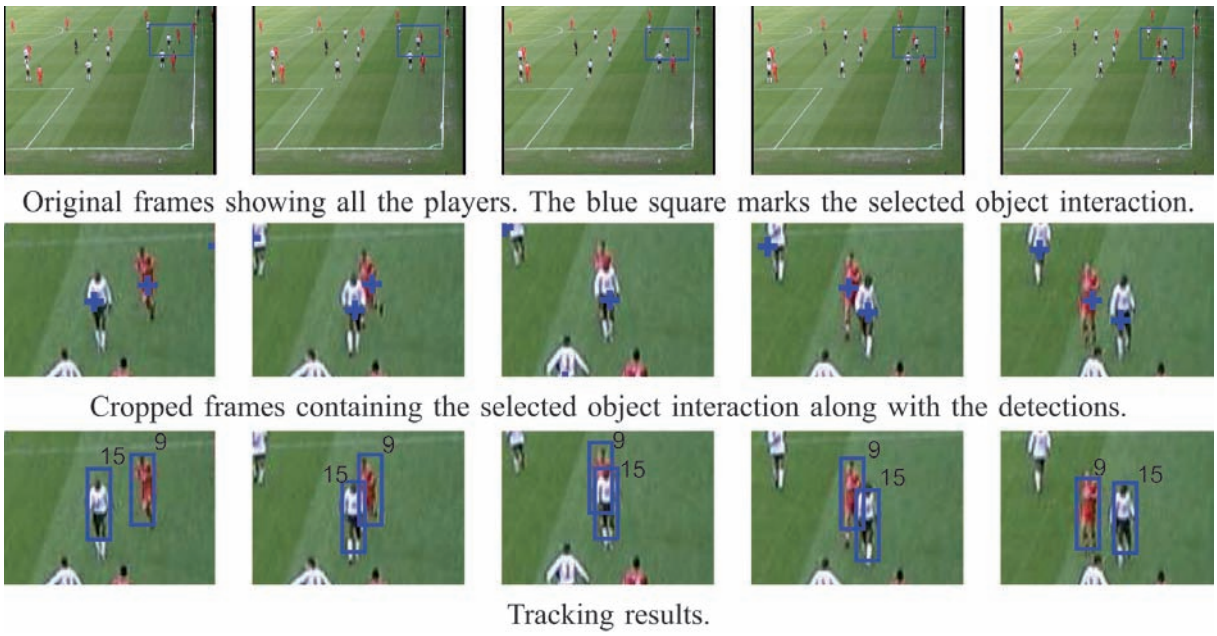

Figure 7 Tracking results for a simple cross between rival players.

in a set of interacting situations extracted from the camera 3 in the 'VS-PETS 2003' dataset. Situations not involving object interactions or occlusions are not considered since they are handled almost perfectly, avoiding in this way that the good results obtained in non-interacting situations obscure the real performance in interacting ones. A tracking error is considered to occur when the distance between the object positions of the estimation and the ground truth is greater than a specific threshold determined by the object size. There is no tracking reinitialization in the case of tracking failure, which allows to test the failure recovery capability of the considered techniques.

The results show that the proposed algorithm clearly outperforms the RBMCDA method in complex crosses, which are the most challenging interactions. The reason is that the RBMCDA method cannot handle trajectory changes during occlusions, since it assumes that the involved objects keep invariable their trajectories. On the other hand, the proposed IRBMCDA method explicitly considers this situation computing several object behavior hypotheses. In overtaking actions, the performance of the proposed method is slightly better, and the improvement is more noticeable when the duration of the interaction increases or the object velocities vary during the occlusion. In simple crosses, both algorithms

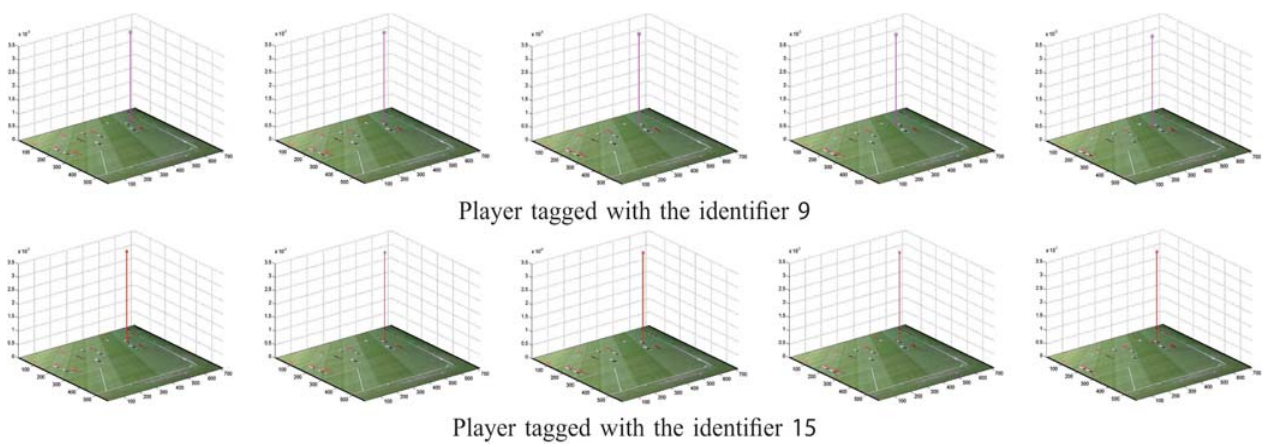

Figure 8 Marginal posterior pdfs of the player trajectories involved in the simple cross of Figure 7.

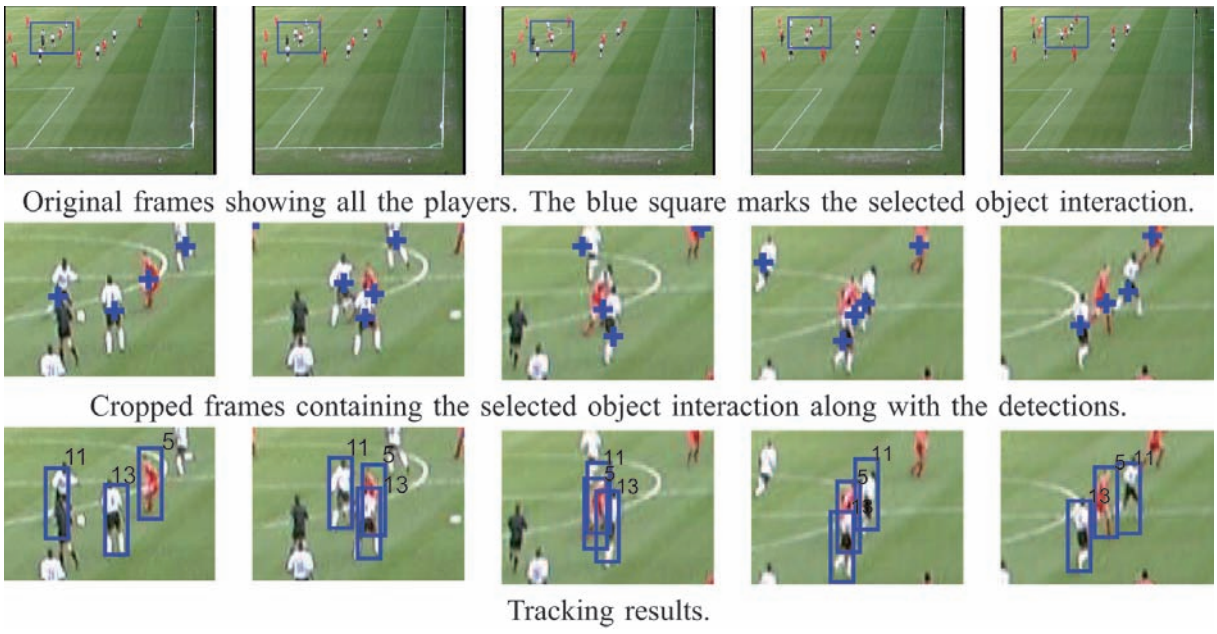

Figure 9 Tracking results for a complex cross involving three players.

correctly estimate the object trajectories since there are no changes in the object trajectories.

The main source of errors arises from situations involving players of the same team, since there is not enough information to reliably estimate the data association. A more sophisticated object detector would be needed, which provides richer information such as pose and shape. In spite of this fact, the tracking algorithm is able to identify when the trajectory estimation is not very reliable, since its variance is significantly higher in these cases.

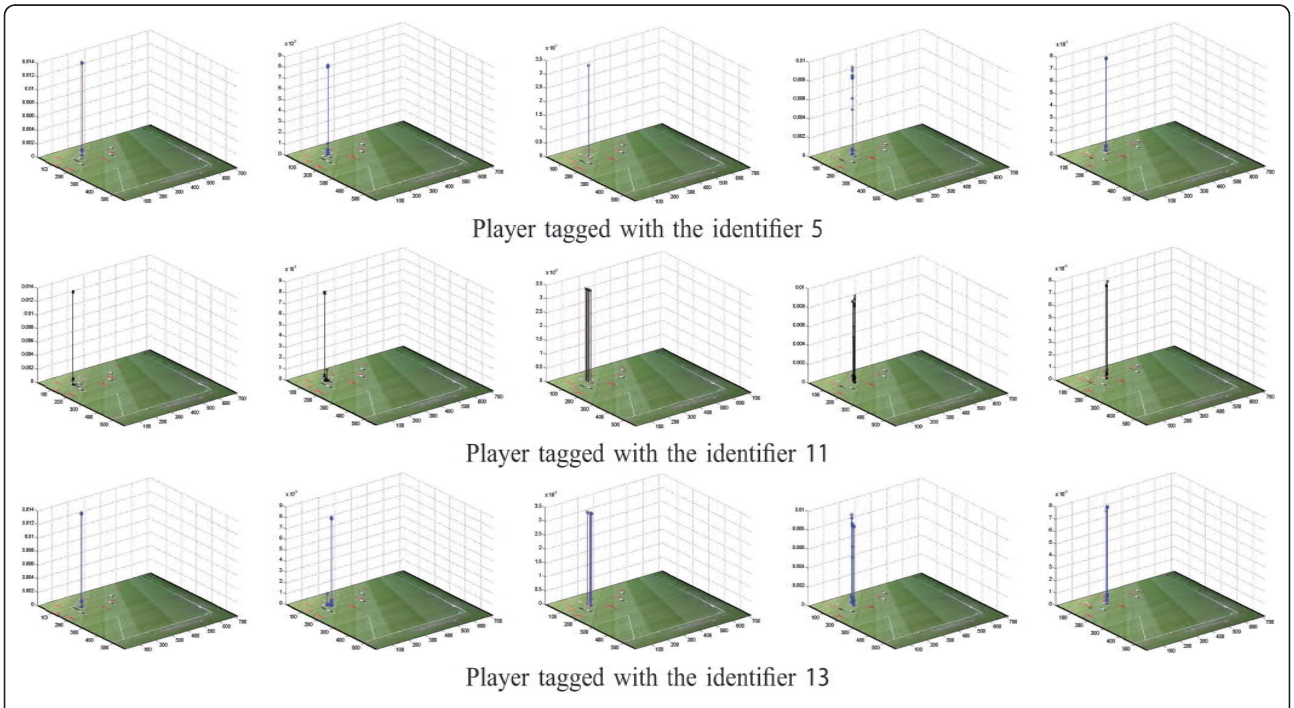

Figure 10 Marginal posterior pdfs of the player trajectories involved in the complex cross of Figure 9.

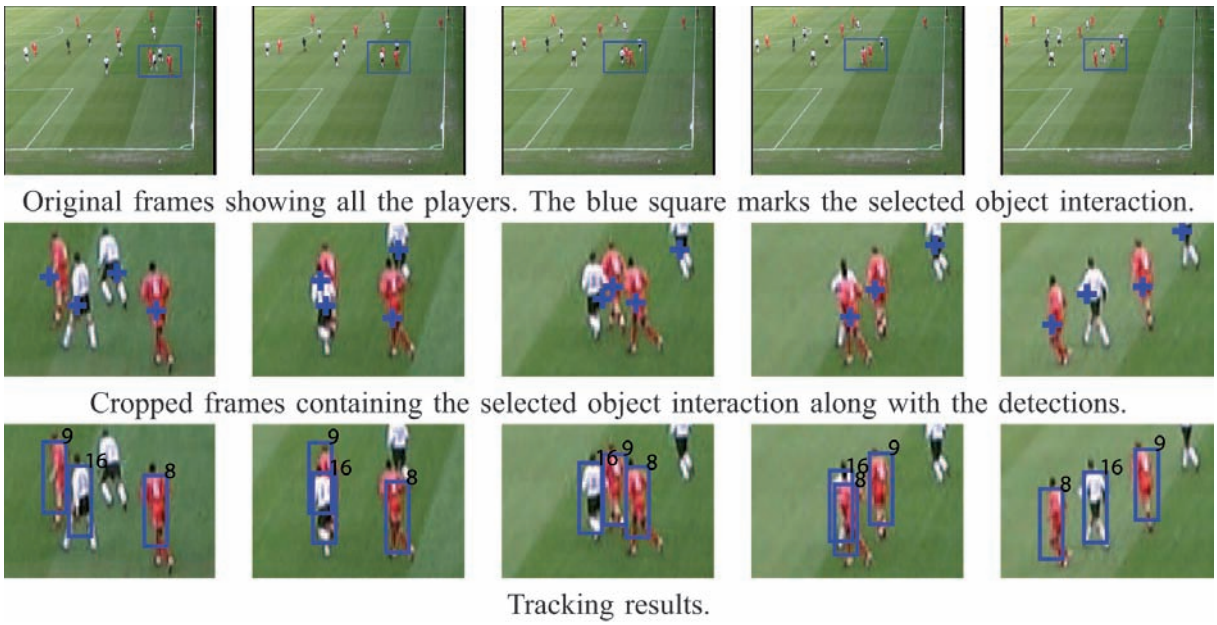

Figure 11 Tracking results for overtaking action involving three players.

8 Conclusions

A novel Bayesian tracking model for interacting objects has been presented. One of the main contribution is an object dynamic model that is able to simulate the object interactions using the predicted occlusion events among objects. The tracking algorithm is also able to handle false and missing detections through a probabilistic data association stage. For the inference of object trajectories, a Rao-Blackwellized particle filtering technique has been used, which is able to obtain accurate estimations in the presence of a high number of tracked objects. In addition, the presented tracking model can work with any

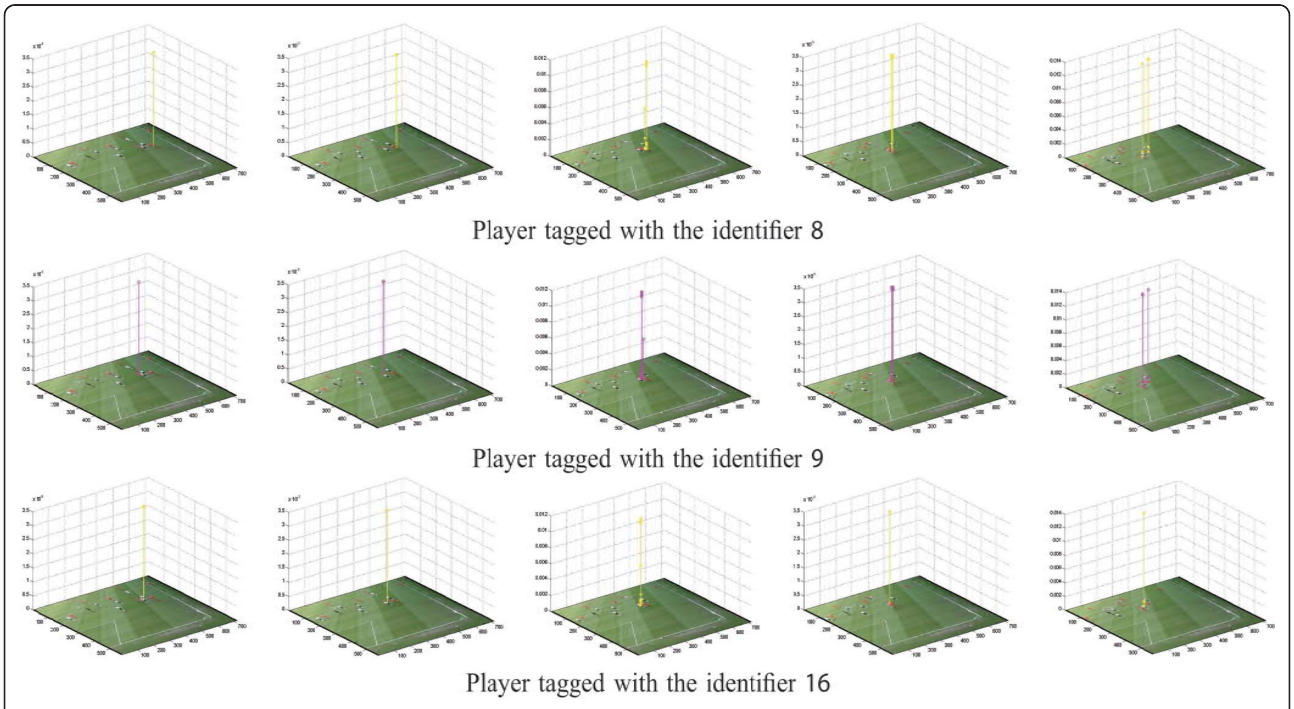

Figure 12 Marginal posterior pdfs of the player trajectories involved in the overtaking action of Figure 11.

Table 1 Tracking results for the proposed IRBMCDA algorithm and the RBMCDA algorithm used for comparison purposes

Object interaction description					Tracking results	
Interaction name	**Interaction type**	**Number of players in interaction**	**Total number of players**	**Duration of interaction in frames**	**Number of errors for IRBMCDA method(the proposed one)**	**Number of errors for RBMCDA method**
interact-1	Simple cross	2	17	46	0	0
interact-2	Simple cross	3	17	72	0	0
interact-3	Simple cross	2	18	48	0	0
interact-4	Simple cross	2	18	50	0	0
interact-5	Simple cross	2	17	123	0	0
interact-6	Simple cross	2	16	99	0	0
interact-7	Simple cross	2	5	37	0	0
interact-8	Simple cross	2	5	56	0	0
interact-9	Simple cross	2	18	73	0	0
interact-10	Complex cross	2	14	36	0	0
interact-11	Complex cross	3	14	56	0	0
interact-12	Complex cross	2	13	55	3	36
interact-13	Complex cross	3	17	78	0	45
interact-14	Complex cross	2	15	69	0	0
interact-15	Complex cross	2	18	61	0	0
interact-16	Complex cross	2	17	113	0	87
interact-17	Complex cross	2	16	109	0	74
interact-18	Complex cross	2	17	50	0	0
interact-19	Complex cross	2	8	92	0	47
interact-20	Complex cross	2	10	126	0	84
interact-21	Complex cross	3	16	45	6	32
interact-22	Complex cross	2	18	38	0	0
interact-23	Overtaking	2	17	95	0	0
interact-24	Overtaking	2	17	60	0	0
interact-25	Overtaking	3	14	94	0	0
interact-26	Overtaking	2	14	35	13	14
interact-27	Overtaking	3	19	89	0	0
interact-28	Overtaking	2	19	29	12	15
interact-29	Overtaking	2	17	108	0	0
interact-30	Overtaking	2	15	90	0	0
interact-31	Overtaking	2	15	89	0	0
interact-32	Overtaking	2	10	27	1	2
interact-33	Overtaking	2	8	63	0	0
interact-35	Overtaking	2	14	100	0	0
interact-36	Overtaking	2	16	45	14	16

object detector that provides at least positional information. The performed experiments have shown a great efficiency and reliability, especially in situations involving complex object interactions where the objects change their trajectories while they are occluded.

Acknowledgements
This study has been partially supported by the Ministerio de Ciencia e Innovación of the Spanish Government under the Project TEC2010-20412 (Enhanced 3DTV).

Competing interests
The authors declare that they have no competing interests.

Received: 14 May 2011 Accepted: 12 December 2011
Published: 12 December 2011

References

1. RE Bellman, *Dynamic Programming* (Courier Dover Publications, New York, 2003)
2. A Doucet, Nd Freitas, KP Murphy, SJ Russell, Rao-Blackwellised particle filtering for dynamic Bayesian networks, in *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 176–183 (2000)
3. S Blackman, *Multiple-target Tracking with Radar Applications* (Artech House, Dedham, 1986)
4. D Reid, An algorithm for tracking multiple targets. IEEE Trans Automat Control. **24**(6), 843–854 (1979). doi:10.1109/TAC.1979.1102177
5. S Blackman, Multiple hypothesis tracking for multiple target tracking. IEEE Trans Aerospace Electronic Syst Mag. **19**(1), 5–18 (2004)
6. IJ Cox, A review of statistical data association for motion correspondence. Int J Comput Vis. **10**(1), 53–66 (1993). doi:10.1007/BF01440847
7. T Fortmann, Y Bar-Shalom, M Scheffe, Sonar tracking of multiple targets using joint probabilistic data association. IEEE J Oceanic Eng. **8**(3), 173–184 (1983). doi:10.1109/JOE.1983.1145560
8. LY Pao, Multisensor multitarget mixture reduction algorithms for tracking. J Guidance Control Dynamics **17**, 1205–1211 (1994). doi:10.2514/3.21334
9. D Salmond, Mixture reduction algorithms for target tracking in clutter, in *SPIE Signal and Data Processing of Small Targets 1990*, **1305**(1), 434–445 (1990)
10. H Gauvrit, J Le Cadre, A formulation of multitarget tracking as an incomplete data problem. IEEE Trans Aerospace Electronic Syst. **33**, 1242–1257 (1997)
11. R Streit, T Luginbuhl, Maximum likelihood method for probabilistic multi-hypothesis tracking, in *SPIE Proceedings of the Signal and Data Processing of Small Targets*. **2235**, 394–405 (1994)
12. C Hue, J Le Cadre, P Perez, Tracking multiple objects with particle filtering. IEEE Trans Aerospace Electronic Syst. **38**(3), 791–812 (2002). doi:10.1109/TAES.2002.1039400
13. Z Khan, T Balch, F Dellaert, Mcmc-based particle filtering for tracking a variable number of interacting targets. IEEE Trans Pattern Anal Mach Intell. **27**, 1805–1918 (2005)
14. CR del Blanco, F Jaureguizar, N García, Robust tracking in aerial imagery based on an ego-motion Bayesian model. EURASIP J Adv Signal Process. **2010**(30), 1–18 (2010)
15. N Gordon, A Doucet, Sequential Monte Carlo for maneuvering target tracking in clutter, in *SPIE Proceedings of the Signal and Data Processing of Small Targets*. **3809**, 493–500 (1999)
16. A Doucet, B Vo, C Andrieu, M Davy, Particle filtering for multi-target tracking and sensor management, in *Proceedings of the International Conference on Information Fusion*. **1**, 474–481 (2002)
17. C Cuevas, CR del Blanco, N Garcia, F Jaureguizar, Segmentation-tracking feedback approach for high-performance video surveillance applications. in *IEEE Proceedings of the Southwest Symposium on Image Analysis Interpretation*, 41–44 (2010)
18. S Särkkä, A Vehtari, J Lampinen, Rao-Blackwellized particle filter for multiple target tracking. J Inf Fusion **8**(1), 2–15 (2007). doi:10.1016/j.inffus.2005.09.009
19. E Maggio, M Taj, A Cavallaro, Efficient multitarget visual tracking using random finite sets. IEEE Trans Circuits Syst Video Technol. **18**(8), 1016–1027 (2008)
20. R Mahler, Phd filters of higher order in target number. IEEE Trans Aerospace Electronic Syst. **43**(4), 1523–1543 (2007)
21. B-N Vo, B-T Vo, N-T Pham, D Suter, Joint detection and estimation of multiple objects from image observations. IEEE Trans Signal Process. **58**(10), 5129–5141 (2010)
22. G Pulford, Taxonomy of multiple target tracking methods, in *IEE Proceedings of the Radar Sonar and Navigation*. **152**(5), 291–304 (2005). doi:10.1049/ip-rsn:20045064
23. Y Ma, Q Yu, I Cohen, Target tracking with incomplete detection. Comput Vision Image Understanding. **113**(4), 580–587 (2009). doi:10.1016/j.cviu.2009.01.002
24. Z Khan, T Balch, F Dellaert, Multitarget tracking with split and merged measurements. in *IEEE Proceedings of the Conference on Computer Vision and Pattern Recognition*. **1**, 605–610 (2005)
25. M Piccardi, Background subtraction techniques: a review, in *IEEE Proceedings of the International Conference on Systems, Man and Cybernetics*. **4**, 3099–3104 (2004)
26. CR del Blanco, F Jaureguizar, N Garcia, Visual tracking of multiple interacting objects through Rao-Blackwellized data association particle filtering, in *IEEE Proceedings of the International Conference on Image Processing*, 821–824 (2010)
27. CM Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer, Berlin, 2006)
28. S Lauritzen, *Graphical Models*, 1st edn. (Clarendon Press, Oxford, 1996)
29. S Arulampalam, S Maskell, N Gordon, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans Signal Process. **50**, 174–188 (2002). doi:10.1109/78.978374
30. D MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, Cambridge, 2003)
31. PI INMOVE (2003) Vs-pets 2003 [Online]. Available: http://www.cvg.cs.rdg.ac.uk/VSPETS/vspets-db.html

doi:10.1186/1687-6180-2011-130
Cite this article as: del Blanco *et al.*: An advanced Bayesian model for the visual tracking of multiple interacting objects. *EURASIP Journal on Advances in Signal Processing* 2011 **2011**:130.